

// **Module 3 Final Project**

Corroborating and Anticipating Drug Use

Kevin Evans
&
Jonathan Vasquez

Contents

- Overall Goal
- Data Gathering / Packages Used
- Exploratory Data Analysis (EDA) / Hypothesis Testing
- Classification Models Created
- Logistic Regression Model Data
- Conclusion
- Future work

// Overall Goal

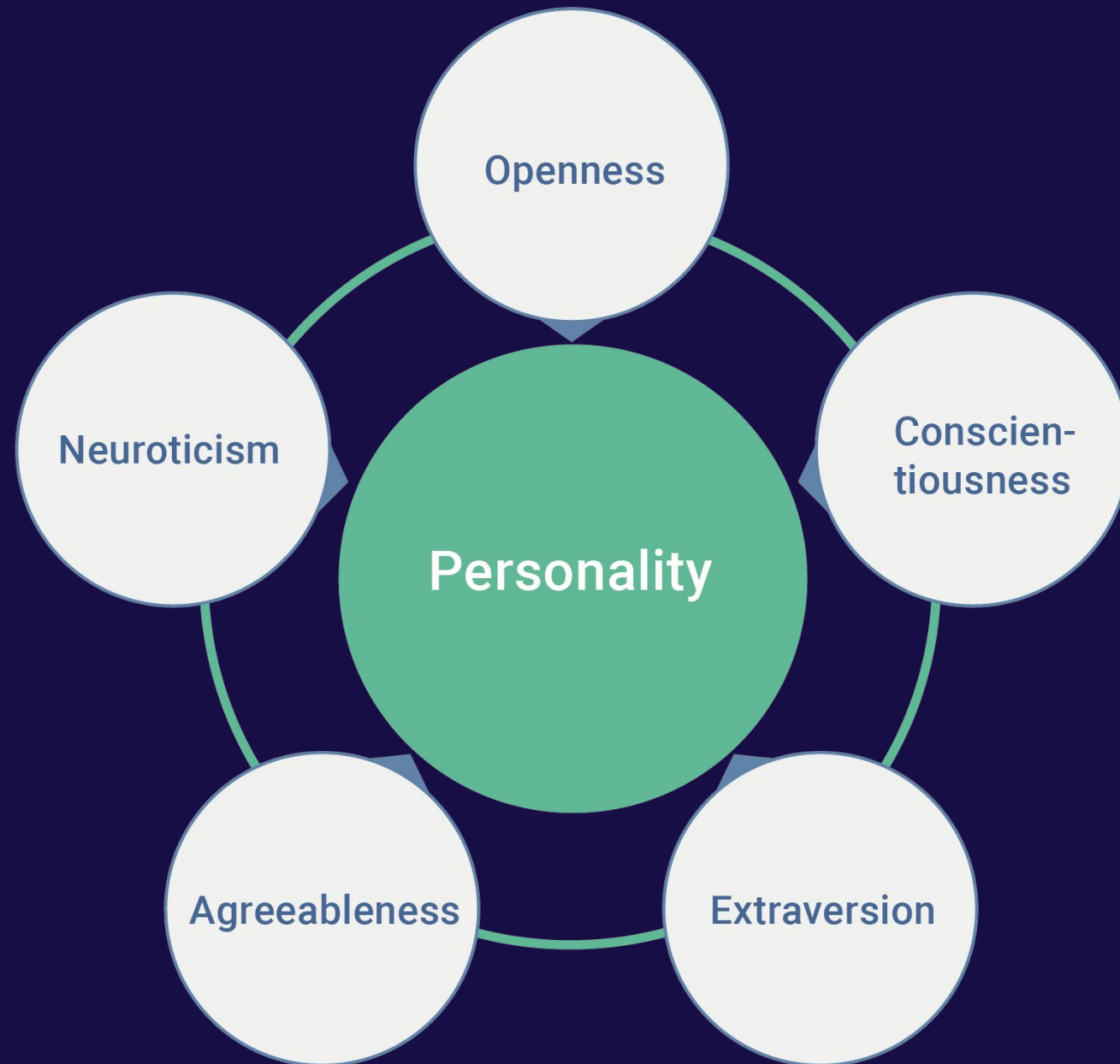
Predict “Hard Drug Users” via:

- 7 Personality Traits:
 - Impulsiveness, Sensation Seeking, & ----->
- 5 Demographic Categories:
 - Age, Gender, Education, Country, & Ethnicity


What’s a “Hard Drug User”?

- A Cocaine, Crack, Heroin, & Meth user within the last decade.

Best Accuracy: Logistic
Regression ~72.94%



// Data Gathering / Packages Used



UCI
Machine Learning Repository
Center for Machine Learning and Intelligent Systems

Drug consumption (quantified) Data Set

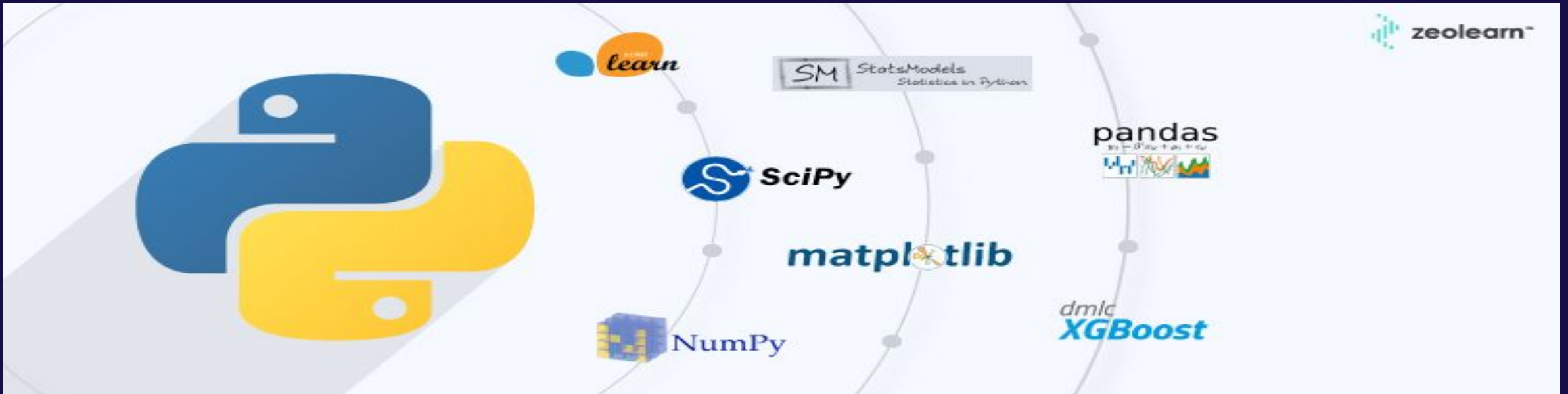
Data Gathering:

Source: [UCI MLR Drug Consumption](#)

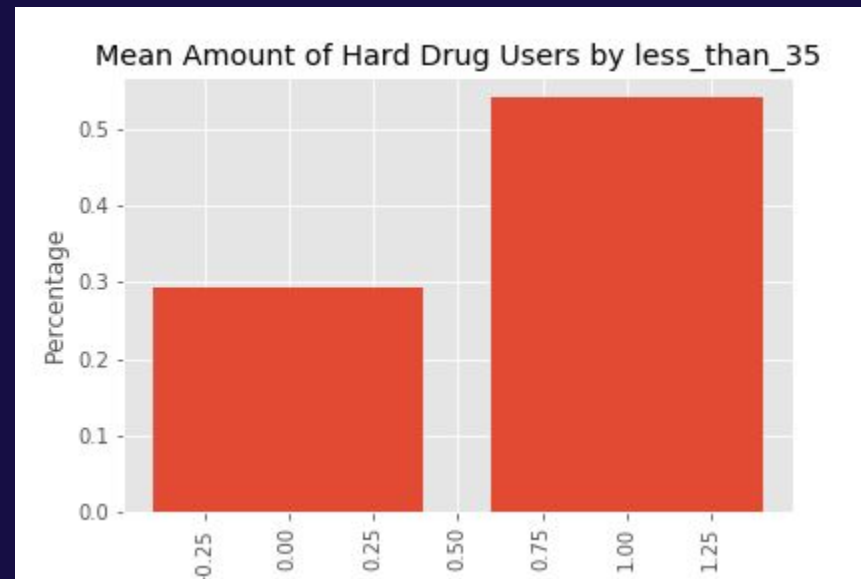
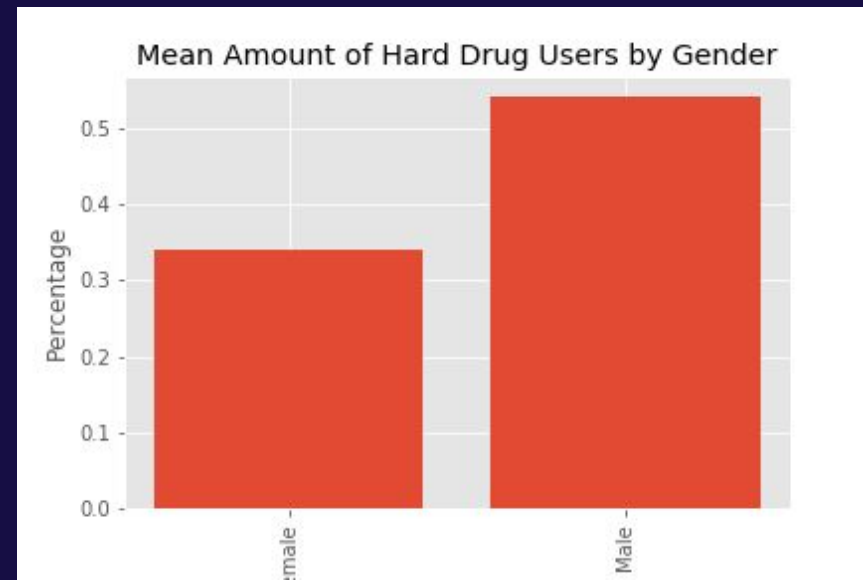
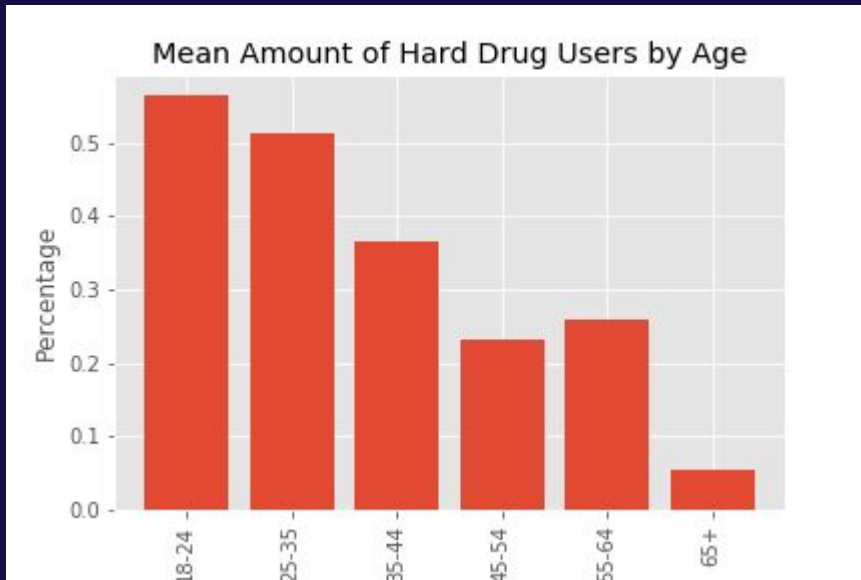
Collaboration: Rampton Hospital,
University of Nottingham, & University of
Leicester

Break-down: 1885 respondents & 12
attributes

Year: 2016



// EDA / Hypothesis Testing



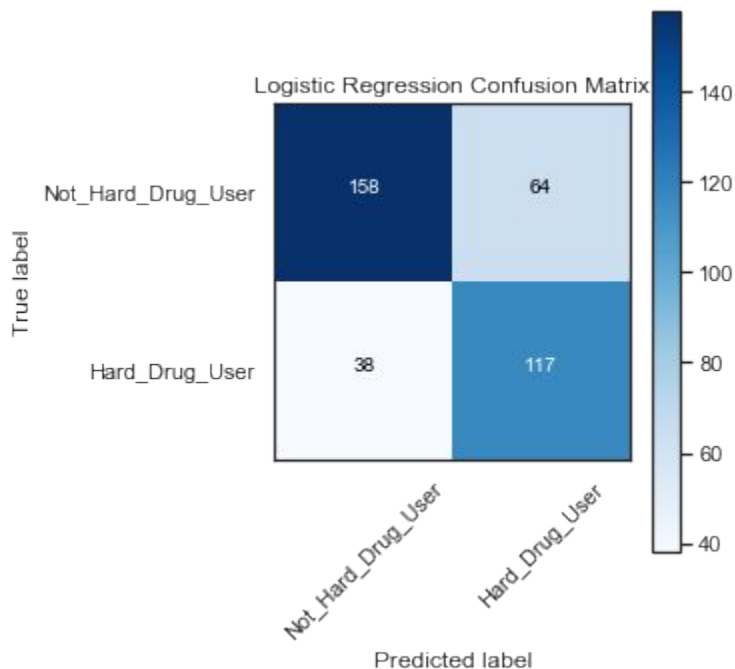
// Classification Models Created

Classification	
Predict: Yes/No, Class	
Logistic Regression	72.94%
Decision Trees	64.19%
Random Forests	72.94%
Adaboost	72.15%
Xgboost	71.35%
KNN	68.70%

- Despite using Grid Search Cross Validation, default Logistic Regression generated highest accuracy on testing data.
- Decision Tree not a strong model for our data.
- Random Forest tied as highest scoring model, but slightly higher FP/FN.
- AdaBoost/XGBoost performed well.
- K Nearest Neighbor model, not a good model to produce prediction.

- Went with Logistic Regression model for final as it has the added benefit of interpretability.

// Logistic Regression Model Data



Nscore	Escore	Oscore	Ascore	Cscore	Impulsive	SS	Less than 35
0.523651	0.49763	0.561872	0.443172	0.431918	0.511018	0.614236	0.519912

Left school at 16 years	Left school at 17 years	Left school at 18 years	Left school before 16 years	Masters Degree	Professional certificate or diploma	Some college or university	University degree
0.520217	0.519116	0.50933	0.51512	0.481408	0.526863	0.518791	0.514023

25-35	35-44	45-54	55-64	Age_65+	Male	Canada	New Zealand
0.583794	0.532198	0.467807	0.471434	0.452885	0.554591	0.54135	0.490198

Other Country	Ireland	UK	USA	Black	Mixed-Black/ Asian	Mixed-White/ Asian	Mixed-White/ Black
0.487432	0.495475	0.455479	0.618778	0.495924	0.502118	0.544689	0.559691

Other Ethnicity	White
0.56886	0.620232

// Conclusion

- Logistic Regression most successful & interpretable model
- Accuracy score ~72.94%
- Highest coefficient probabilities:
 - Ethnicity: White, 0.62032
 - Country: USA, 0.618778
 - Personality Trait: SS & Oscore, 0.614236 & 0.561872
 - Age: 25-35, 0.583794

// Future work

- Create different classification of drugs; party, hallucinogens, etc.
- Create different separation in times, our break was by decade, but it could have been year, month, or week.
- Devise a plan that can be a proactive measure in alleviating government resources spent on “Hard Drug Users”. As per, NIH/NIDA the US spends \$193 Billion.

Questions?



Thank You!

Kevin Evans &
Jonathan Vasquez